Journal of Volcanology and Geothermal Research 381 (2019) 44–56

Contents lists available at ScienceDirect

Journal of Volcanology and Geothermal Research

journal homepage: www.elsevier.com/locate/jvolgeores

Forecasting explosion repose intervals with a non-parametric Bayesian survival model: Application to Sakura-jima volcano, Japan

S.F. Jenkins [a,b,*], H. Goldstein [c], M.S. Bebbington [d], R.S.J. Sparks [a], Takehiro Koyaguchi [e]

[a] School of Earth Sciences, University of Bristol, Bristol, United Kingdom
[b] Earth Observatory Singapore, Asian School of the Environment, Nanyang Technological University, Singapore
[c] Centre for Multilevel Modelling, University of Bristol, Bristol, United Kingdom
[d] Volcanic Risk Solutions, Massey University, New Zealand
[e] Earthquake Research Institute, University of Tokyo, Yayoi, Tokyo, 113, Japan

ARTICLE INFO

Article history:
Received 21 December 2017
Received in revised form 17 April 2019
Accepted 17 April 2019
Available online 22 April 2019

Keywords:
Vulcanian explosions
Repose
Bayesian Markov Chain Monte Carlo modelling
Sakura-jima

ABSTRACT

Forecasting the repose between eruptions at a volcano is a key goal of volcanology for emergency planning and preparedness. Previous studies have used the statistical distribution of prior repose intervals to estimate the probability of a certain repose interval occurring in the future, and to offer insights into the underlying physical processes that govern eruption frequency. However, distributions are only decipherable after the eruption, when a full dataset is available, or not at all in the case of an incomplete time-series. Thus there is value in using an approach that does not assume an underlying distribution in forecasting likely repose intervals, and that can make use of additional information that may be related to the duration of repose. The use of a non-parametric survival model is novel in volcanology, as the size of eruption records is typically insufficient. Here, we apply a non-parametric Bayesian grouped time Markov Chain Monte Carlo (MCMC) survival model to the extensive 58-year eruption record (1956 to 2013) of Vulcanian explosions at Sakura-jima volcano, Japan. The model allows for the use of multiple observed and recorded data sets, such as plume height or seismic amplitude, even if some of the information is incomplete. Thus any relationships between explosion variables and subsequent or prior repose interval can be investigated. The model was successfully able to forecast future repose intervals for Sakura-jima using information about the prior plume height, plume colour and repose durations. For plume height, smaller plumes are followed by shorter repose intervals. This provides one of the first statistical models that uses plume height to quantitatively forecast explosion frequency.

1. Introduction

Knowledge of the likely repose between eruptions at a volcano can inform emergency management actions and offers insights into the underlying physical processes that govern eruption frequency. Many studies have attempted to understand eruption repose intervals over timescales that vary from hours or days (e.g. Watt et al., 2007; Connor et al., 2003; Jaquet et al., 2006) to decades or centuries (e.g. Pyle, 1998; Mendoza-Rosas and De La Cruz-Reyna, 2010), or both (e.g. Marzocchi and Bebbington, 2012). A common model is a renewal process (Bebbington and Lai, 1996), in which the repose interval duration depends only on the time since the last onset, commonly inferred to be due to magma recharge (Marzocchi and Zaccarelli, 2006; Turner et al., 2011). A mixture of the Weibull renewal model was proposed by Turner et al. (2008) to represent the effects of mantle recharge into a volcano's plumbing system. At Sakura-jima, Udagawa et al. (1999) found that a lognormal probability distribution best fit the 1980–1994 time-series of explosion repose intervals and suggest that the volcanic system is characterised by branching conduits that control the renewal process. Other processes have been inferred from the distribution of repose intervals for open-system volcanoes with frequent Vulcanian explosions, including Sakura-Jima. Log-logistic (Connor et al., 2003) and Weibull (Watt et al., 2007) distributions were used to signify, respectively, competing processes in the conduit and a classic failure model.

Statistical models that forecast eruption onsets need to be fitted to the prior repose interval data. Which of multiple candidates best fit the data can be identified by means of maximum likelihood-based techniques. A more difficult question is whether any of the models capture the essential features of the data. This question is normally answered via residual analysis, which in the case of renewal models can be based on the point-process compensator (Ogata, 1988). This approach provides formal tests where the rescaled process can be tested against a null hypothesis of a Poisson process. Given a large enough dataset, not something commonly available for eruptions, an alternative is to

* Corresponding author.
E-mail address: susanna.jenkins@gmail.com (S.F. Jenkins).

https://doi.org/10.1016/j.jvolgeores.2019.04.008

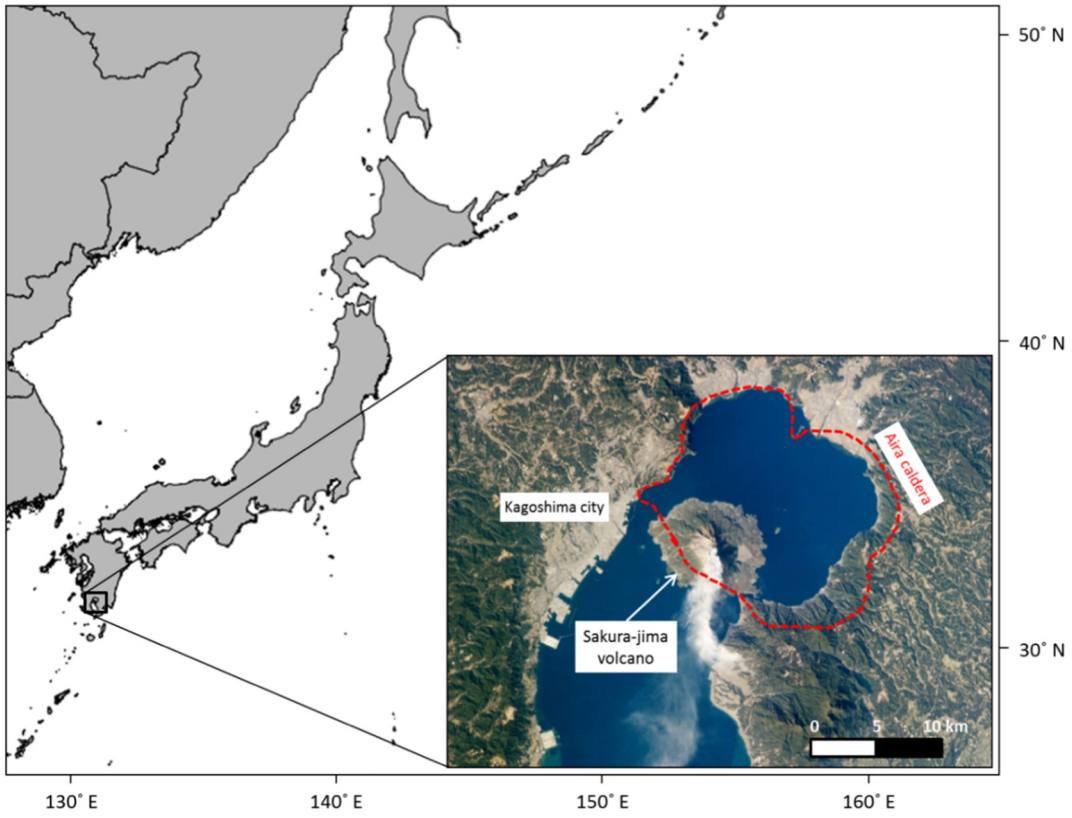

Fig. 1. Location of Sakura-jima volcano, Aira caldera and Kagoshima city within Japan. NASA Satellite image taken from the International Space Station, 10 January 2013.

test the forecasts from the model against future data not used in the model development or fitting (Bebbington, 2013a). These methods typically work well for datasets up to a few hundred observations. Beyond this, simple parametric renewal models with handfuls of parameters are typically overwhelmed by the complexities that then become apparent in the data. The tendency towards complexity with larger datasets can be countered by division of the dataset into different periods, basically cutting down the size of the dataset to something manageable, or by a non-parametric approach as used here, in which the data are used directly in the model formulation.

For some volcanoes, including Sakura-Jima, there are data available beyond simply the onset times, and therefore repose durations, such as a measure of the eruption size (Marzocchi and Zaccarelli, 2006), or the geochemistry of prior deposits (Green et al., 2013). This additional information can be incorporated into a statistical renewal model through a proportional hazards approach, as we do here. In this study, we investigate the use of a non-parametric Bayesian grouped time Markov Chain Monte Carlo (MCMC) survival model in forecasting explosion repose intervals at Sakura-jima volcano in Japan. In comparison to previous statistical distribution fits to the data, the advantage of the model proposed here as a tool for forecasting future repose durations is that: 1) no particular underlying distribution to the data is assumed; 2) explosion variables, such as plume height and size, can be incorporated into model forecasts; and 3) a Bayesian multiple imputation technique (described in Section 4) is used to deal with missing data, thereby increasing the number of usable records.

In what follows, we introduce Sakura-jima volcano and the eruption dataset (Section 2) before applying parametric statistical models to the data (Section 3), which provide rationale for the use of the non-parametric model described in Section 4 and fitted, validated and applied to forecast repose durations in Sections 5 and 6. Section 7 discusses the model application and draws relevant conclusions for our study.

2. The Sakura-jima eruption: 1956 to 2013

Sakura-jima volcano is an andesitic stratovolcano located in Kagoshima Bay, southern Kyushu, approximately 4 km from Kagoshima city. The volcano is part of a long-lived volcanic system associated with several nested caldera structures, of which Aira caldera is the youngest, forming 22,000 years ago (Aramaki, 1984) (Fig. 1). The post-caldera phase from 13 ka to present has consisted of major explosive eruptions (most recently in 1914) with intervening periods of prolonged Vulcanian explosions and ash-venting over decades to centuries (Kobayashi et al., 1988). Most recently, the volcano has been active since 1955 and regularly produces small (<3 km plume height) Vulcanian explosions. Explosions have been continuously monitored by the Kagoshima Local Meteorological Observatory of the Japan Meteorological Agency (JMA), with records including time of explosion, plume height, a proxy for size[1] and colour, and maximum seismic and infrasonic amplitude, which we term 'explosion variables'. The record represents the longest continuous dataset of Vulcanian explosions and ash emissions available. Each explosion is short-lived (order of a few minutes or less) with short impulsive seismic signals. Between 1955 and 2006, explosions were produced exclusively from one or both of the summit 'Minamidake' craters. In June 2006, the 'Showa' crater, approximately 500 m to the East of Minamidake, also became active (Yokoo et al., 2013) and by 2009 was the source for the vast majority of explosions.

Details of the Sakura-jima eruption dataset of explosion times and variables compiled for this study are given in Table 1. Two different

[1] The plume size (Q) is an integer value derived from the vertical cross-section (S) in m^2 through the function $Q = (\log_{10}S - 4.328) / 0.358$, when areas are between 10^4.865 and 10^6.297. The decimal value is rounded to the nearest integer to provide a proxy for plume size. Areas below 10^4.865 are assigned Q = 1, and above 10^6.297 assigned Q = 6. Where columns reach to the stratosphere Q = 7.

Table 1
Description of data available for explosions measured throughout the Sakura-jima eruption record from 1956 to 2013. Additional data include qualitative observations of explosive sounds, sensory infrasonic wave, volcanic projectiles, thunder and changes in pressure as well as the direction of ash transport. Information on seismic and infrasonic amplitudes are easily accessible up to and including the year 2000, after which the record format changed so that comparison on an event-by-event base is difficult. The derivation of the plume size proxy is explained in an earlier footnote.

1956 to 2013				
Variable	n	% data available	Minimum	Maximum
Repose interval (hours)	11,828	100	0.02 (1 min)	7394 (308 days)
Plume size proxy (Q)	7151	62	1	6
Plume height (m)	7416	61	50	5000
Plume colour	5972	50	Four key colours: black, grey, brown, white (ten in total)	
Explosion crater	1989	52	One or multiple of three craters	
Maximum seismic amplitude(μm)	6556	55 (1961–2000)	0.4	204.0
Infrasound (hPa) at site 0	1690	14 (1990–2000)	0.01	3.08
Infrasound (hPa) at site E	544	5 (1997–2000)	0.03	1.35

continuous datasets were merged: one from 1956 to 2001 (n = 7498 explosions: Udagawa et al., 1999), and one from 2001 to 2013 (n = 4330 explosions). The datasets were sourced from Bulletins of the Japan Meteorological Agency (JMA, 1956a; JMA, 1956b), and their online catalogue (JMA, 2015). The data examined here include only those ash emissions that are associated with acoustic or seismic signals to prevent airborne ash recorded during scheduled observations being included as additional explosions, rather than as the remnants of a previous explosion. The data required extensive formatting in order to provide one consistent catalogue that could be used for analysis; the dataset is provided as supplementary material. The explosion variable data are shown in Fig. 2.

Four different explosion craters (Showa, Minamidake A, Minamidake B, and simultaneous Minamidake A B), are recorded. Plume colour is recorded as one of ten different variants. Seismic data are only available from 1961 to 2000, and infrasonic data from 1990 to 2000 and 1997 to 2000 for sites O and E, respectively. Plume heights are estimated from visual observations and are therefore likely to vary according to wind conditions at the time of explosion, which affects the angle of viewing and the inclination of any plume. A discrete proxy for plume size, Q, is provided in the JMA catalogue derived from the vertical cross-sectional area of the plume height and binned into qualitative sizes ranging from a 'very small amount' (1) to 'extremely large amount' (6) and plumes that reach into the stratosphere (7), as described in an earlier footnote. No eruptions of size 7 are present in the 1956 to 2013 dataset. The plume height and Q are very strongly positively correlated (Spearman's rank correlation of 0.75, P-value of 0). The seismic dataset available to us between 1961 and 2000 does not distinguish between earthquake types, e.g. hybrid or explosion earthquakes, preventing separate analysis by earthquake type.

The dataset is not complete across all explosion variables for all time, with some periods showing more complete records than others (Fig. 3) Fig. 4 shows the empirical repose distributions separated according to whether the respective explosion variables were observed or missing for the explosion prior to the repose. We see that missing visual (plume colour, height or size, or explosion crater) observations indicate a longer subsequent repose, whereas the effect is reversed for missing instrumental (seismic, infrasound) observations. Hence we conclude that missing data may be informative and need to be incorporated into our models.

We are interested in the potential information that explosion variables provide for forecasting the subsequent repose. In the case of the non-discrete variables, shown in Fig. 5, we can calculate the rank correlation and consequent P-value against a null hypothesis of zero correlation. Missing values were omitted from the calculation along with the corresponding reposes. While there is a significant (positive) correlation between plume height and subsequent repose, there is no indication of any significant correlation between repose duration and seismic or infrasonic measurements. Although the Infrasound measurements from Site E are marginally significant, they cover only 4 years of eruptions, and possibly represent a false positive for the test of significance given that measurements from Site O were not a significant predictor, and the fact that the two infrasound sequences have a correlation of 0.793.

Turning to the discrete variables, Fig. 6 shows the distributions of repose lengths subsequent to each plume colour. There is a great degree of overlap, but using analysis of variance (ANOVA) on the logarithm of the repose length (to remove the skewness), we find that the colours can be grouped as 'Missing' (5853 observations), 'Grey-White/Grey-Brown/Black' (3190 observations) and 'Other' (2779 observations), with no significant differences in the means within each group (P = 0.380 and P = 0.187 for the Grey-White/Grey-Brown/Black and Other combined groups, respectively). The groups are then different with P < 0.0005.

The effect of the explosion crater location is shown in Fig. 7. In a similar manner to the plume colour, we find that the explosion crater can be grouped as 'Showa' and 'Other' (P = 0.384 between groups).

Turning to the proxy for plume size (Q), the effect of which on repose length is shown in Fig. 8, we find that the Q-values can be grouped as Q = {1,2} (1294 observations, P = 0.58), {3} (3419 observations), {4} (1479 observations) and {missing,5,6} (5632 observations, P = 0.38).

Traditionally, models for repose intervals have been based on renewal models, where each repose is independent and identically distributed (Bebbington and Lai, 1996; Bebbington, 2013a). Serial correlation is not usually incorporated: it is assumed that, conditional on any covariates, the interval lengths are independent. If there is reason to believe that there is a 'memory' effect in which processes are capable of remembering the past dynamic conditions (e.g. Jaquet et al., 2006), we can additionally condition on the previous repose interval length, or the previous two or more repose intervals if the dependence is considered to be greater than first order. The autocorrelation function for the logarithm of the repose lengths is shown in Fig. 9a. We see that the reposes are positively correlated. Much of this is due to trend (s) within the sequence, as demonstrated by the much shorter term autocorrelation of the first differences of the log-reposes (Fig. 9b). This indicates that the previous two reposes should be sufficient to condition the process.

To summarise, the variables that seem to have potential for forecasting repose durations are the prior plume height, plume colour (two categories – missing is treated as a reference group), explosion crater ('Showa' or not), plume area (three categories plus reference group), the previous two log-reposes, and something in the way of a time trend (i.e. changes in eruption rate over time).

3. Parametric models

Before introducing the non-parametric Bayesian survival model, we will consider the parametric alternative(s) for benchmarking purposes.

A shorter record (1955–1998) of the explosions at Sakura-jima has been analysed previously by Watt et al. (2007), who found that a log-logistic distribution best described the repose intervals in the earlier

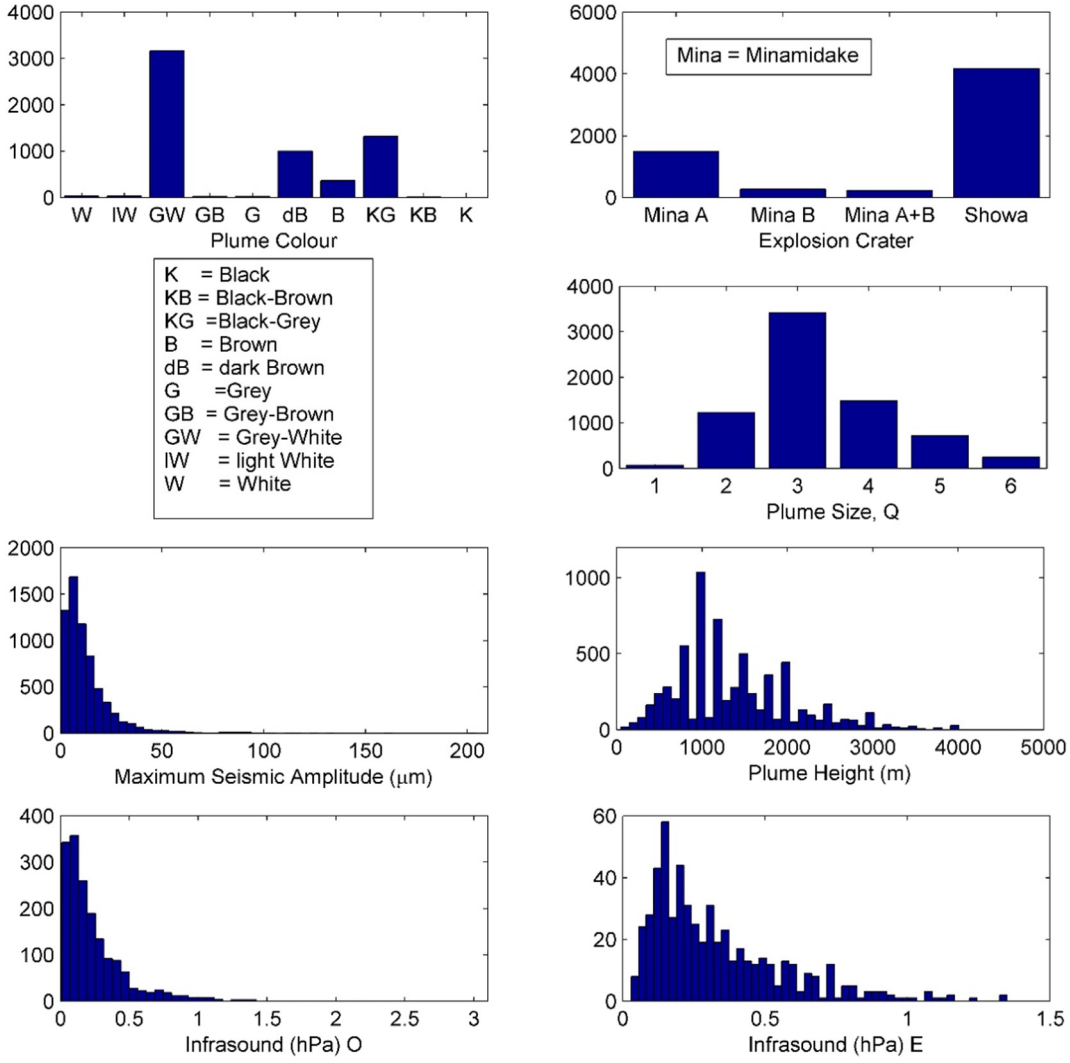

Fig. 2. Data available for explosions measured throughout the Sakura-jima eruption record from 1956 to 2013. Additional data include qualitative observations of explosive sounds, sensory infrasonic wave, volcanic projectiles, thunder and changes in pressure as well as the direction of ash transport.

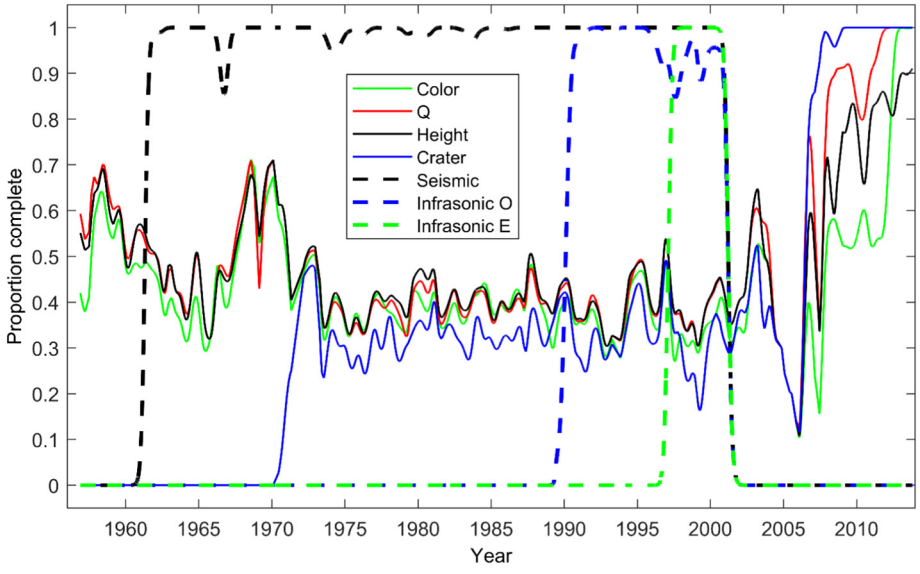

Fig. 3. Completeness over time (Gaussian kernel smoother, bandwidth 100 days) for explosion variables in our dataset.

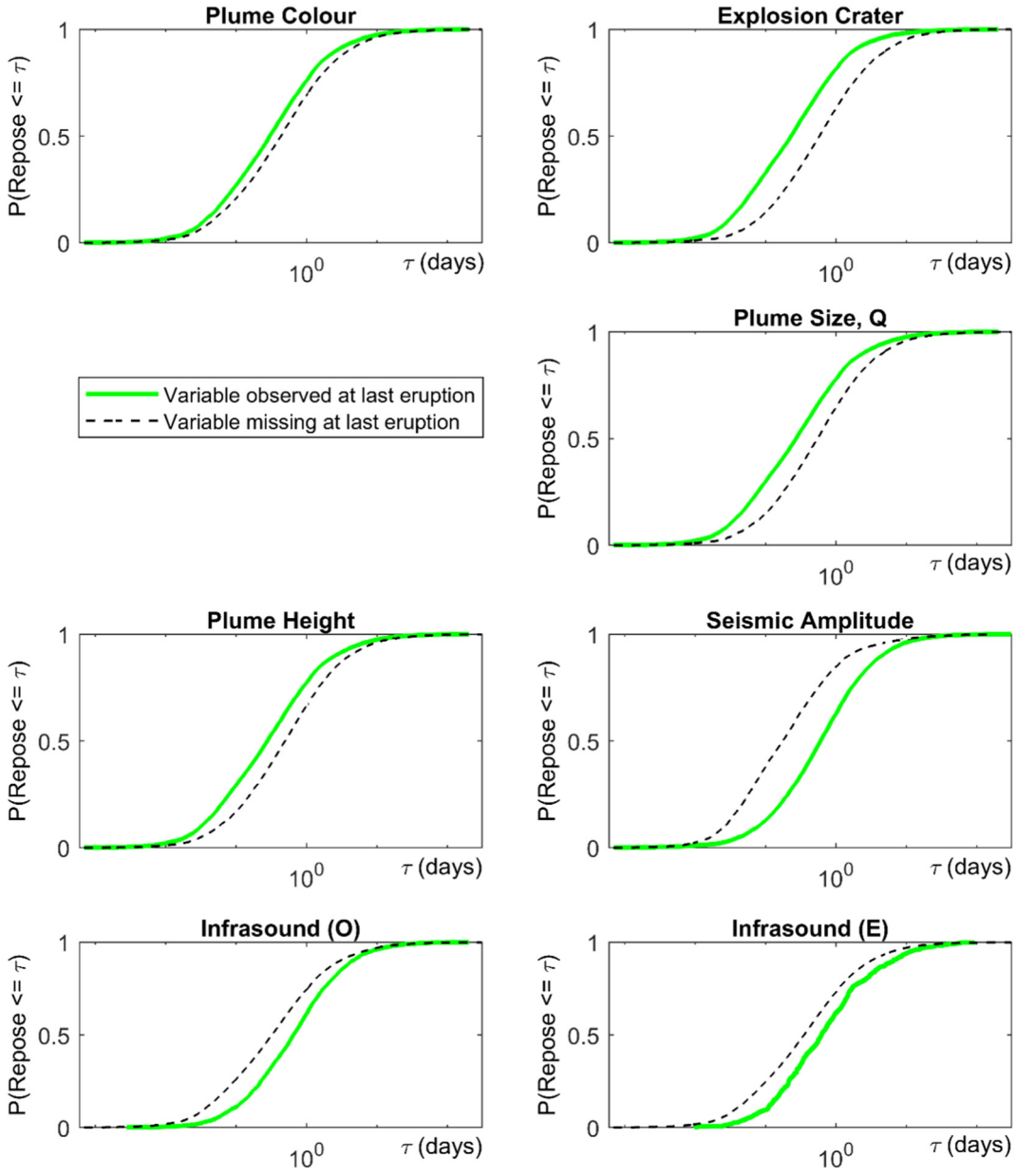

Fig. 4. Repose lengths conditional on observation/non observation of explosion variables.

part of the eruption (up to 1971), and a Weibull distribution the latter part. Note, however, that only the eruption onset times were used to fit the model, not any of the other variables. The fit for Weibull, two and three component mixtures of Weibulls (Turner et al., 2008) and log-logistic renewal distributions to the current dataset are shown in Fig. 10. We see that the log-logistic outperforms all of the Weibull mixture renewal models, as measured by the Bayes Information Criterion, BIC = p log(n) – 2 logL, where p is the number of parameters, n the number of data and L the estimated maximum likelihood. The log-logistic distributions were attributed to explosions being controlled by competing processes in the conduit (Connor et al., 2003; Watt et al., 2007). The log-logistic distribution has probability density

$$f(\tau) = \frac{k\tau^{k-1}}{\mu^k\left(1+(\tau/\mu)^k\right)^2} \quad (1)$$

where τ = t-s is used to denote the elapsed time since the last eruption onset occurred at time s, and k, μ are parameters to be estimated.

The log-logistic k values, which describe the shape of the density function, and therefore the periodicity of explosions, were noticeably lower (<1.3 (Watt et al., 2007), 1.0072 for the current dataset) than those observed at Soufrière Hills Volcano, Montserrat ($k = 4$; Connor et al., 2003). A k value of 1.0 within a log-logistic renewal model would represent Poissonian activity. However, the rate of eruptions varies through time, and successive repose intervals are strongly correlated (Fig. 9), which rules out the Poisson process as a feasible model. Instead we need to investigate the hidden structure controlling the process, developing a model in which information additional to the previously observed repose interval durations is used to improve forecasts of the time to the next eruption. In doing so, we hope to find that the log-logistic model fits the entire data, with appropriate time-varying modulation from the other variables. An alternative could be to try a change-point (or 'regime') type model (Mulargia et al., 1987), but without a stochastic model for future changes in regime this cannot be used for forecasting. Hence it would have to be placed within the framework of a hidden Markov model (Bebbington, 2007).

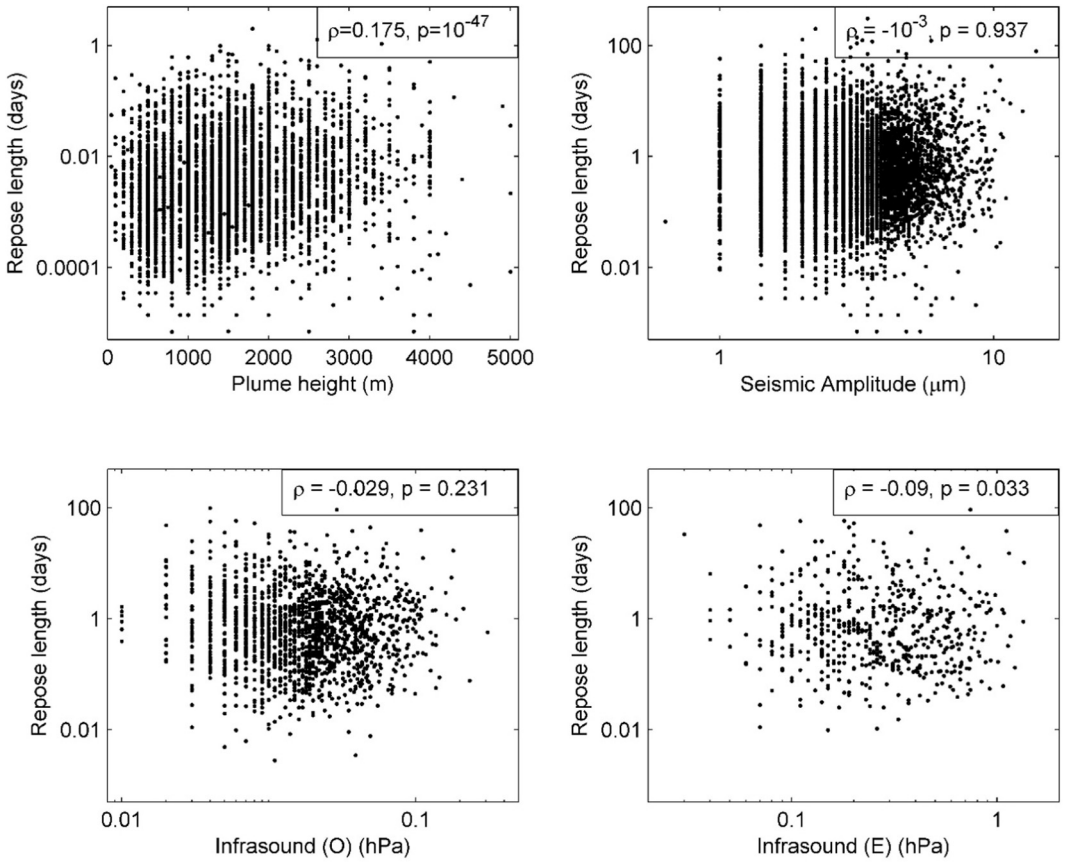

Fig. 5. Rank correlation coefficients (ρ) and P-value (p) between non-discrete explosion variables and subsequent repose. A P-value of close to 0 refutes the null hypothesis of zero correlation.

A possible augmentation to the log-logistic renewal model to account for time-varying behaviour is to formulate it as a proportional hazards model (Green et al., 2013) incorporating eruption variables.

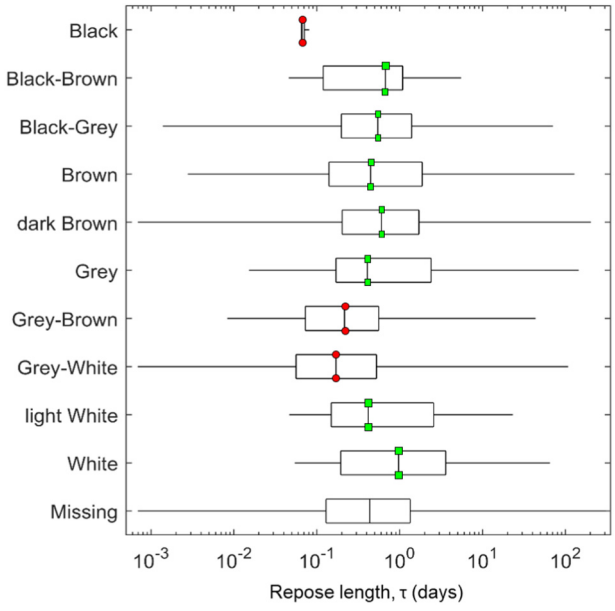

Fig. 6. Repose durations by plume colour. Using ANOVA, plume colours can be grouped as Black, Grey-Brown, Grey-White (red circles), Others (green squares), and Missing (no markers). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

The corresponding survival function for the log-logistic density function of Eq. (1) is defined as

$$S(\tau) = \Pr(\text{Repose length} > \tau) = \frac{\mu^k}{\mu^k + \tau^k},$$

and hence the hazard rate is

$$h(\tau) = f(\tau)/S(\tau) = \frac{k\tau^{k-1}}{\mu^k + \tau^k}$$

where the probability of an onset in a short time interval (τ,τ Δ) is approximately h(τ)Δ. The proportional hazards model is then defined via its hazard function

$$h_{\text{PH}}(\tau) = \exp\left(\delta' x\right) h(\tau) \tag{2}$$

where x is a vector of eruption variables, and δ a vector of coefficients to be determined. From the analysis above, a reasonable starting point for the proportional hazards term is

$$\delta' x = \delta_1 \log h + \delta_2 C_1 + \delta_3 C_2 + \delta_4 S + \delta_5 Q_{12} + \delta_6 Q_3 + \delta_7 Q_4 + \delta_8 \log \tau_{-1} + \delta_9 \log \tau_{-2} + \delta_{10} N(t)/t \tag{3}$$

where h is the height of the previous plume (missing heights will be imputed by the mean), C_i are indicator variables (0 or 1) for the two 'non-missing' plume colours identified (Fig. 6), S is an indicator variable for the Showa crater, Q_{12} is an indicator variable for Q = 1 or 2, similarly for Q_3 and Q_4. Terms 8 and 9 are obtained by noting that there is a strong correlation between $\log\tau_i - \log\tau_{i-1}$ and $\log\tau_{i-1} - \log\tau_{i-2}$, while the last term is an allowance for deviation from the overall rate of events.

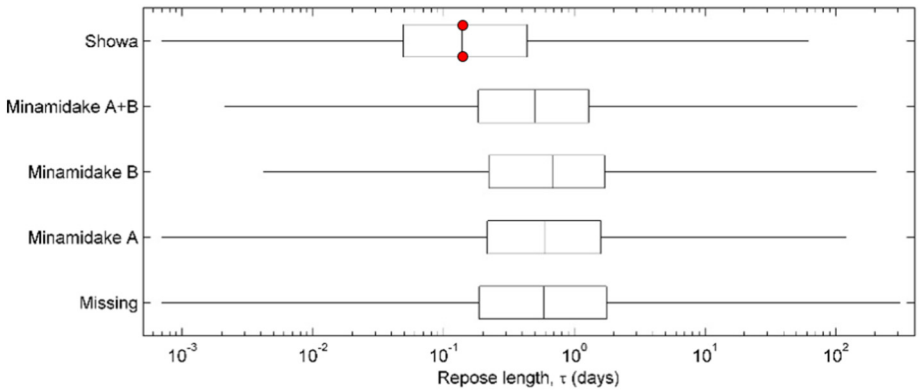

Fig. 7. Repose durations by explosion crater. Using ANOVA, craters can be grouped as Showa (red circles) and 'other' (Minimidake A, B, A B and Missing: no markers). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Note that the proportional hazard is relative to a missing plume colour, from the Minamidake craters, with Q = 5,6 or missing. The continuous measurements are normalized by subtracting their mean.

Assessing the resulting model fits using BIC reveals that the plume colour and plume area (Q) terms are not significant, and hence our baseline model reduces to

$$\delta' x = \delta_1 \log h + \delta_4 S + \delta_8 \log \tau_{-1} + \delta_9 \log \tau_{-2} + \delta_{10} N(t)/t. \quad (4)$$

Evaluation using the point process compensator (Bebbington, 2013b) indicates that the model does not pass the goodness of fit tests (constant rate, independence, exponential distribution). Given these limitations of simple distribution fitting, this study will now focus on investigating the relationship between repose interval and explosion variables without assuming an underlying distribution.

4. The non-parametric Bayesian survival model

The model is an adaptation of a non-parametric model for fitting ordered multilevel data (Goldstein, 2011), available as MATLAB programs (www.bristol.ac.uk/cmm/software/realcom). The basic data structure for the model application here is that of the sequence of (repose) intervals between Vulcanian explosions. Interactions between explosion variables (covariates) are handled within the model so that the more traditional and complex approach illustrated above of arriving at a suitable parametric model is unnecessary. Further details about the derivation of the model and the estimation algorithm are given in Goldstein (2011, Chapter 11). The following description contains the essential concepts and provides the statistical model formulation. We provide a table of symbols and their meanings in Appendix 1.

We adopt a latent normal formulation (Goldstein et al., 2009) for an ordered m-category response, by discretising (or approximating) the repose distribution into a non-parametric piecewise linear survival function. We operationalise the model by discretising the repose length time scale with m cut points ($\tau_0(=0), \tau_1, \ldots, \tau_m$). Thus, if $\varphi(s)$ denotes the standard normal density with zero mean and variance one, we have a probit link cumulative probability model

$$\gamma_h = \Pr(\tau \le \tau_h) = \sum_{g=1}^{h} \pi_g = \int_{-\infty}^{\alpha_h - X'\beta} \varphi(s) ds \quad (5)$$

for time intervals h = 1, …,m-1, where X is a matrix of predictor variables. Eq. (5) is used, given a value of β, to determine defines α_h where β are parameters, treated in the Bayesian formulation as random variables, and $\pi_g = \Pr(\tau_{g-1} < \tau \le \tau_g)$.

In practice, the number and location of cut points, m, will be determined by the data. A useful procedure is to aim for approximately

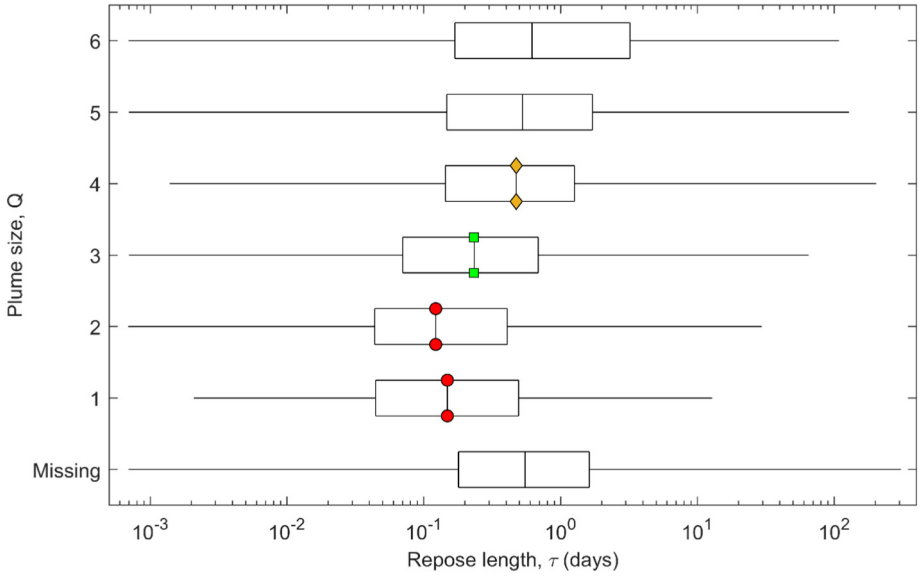

Fig. 8. Repose durations by Q, a proxy for plume size. Using ANOVA, plume sizes can be grouped as 1, 2 (red circles), 3 (green squares), 4 (orange diamonds), and Missing, 5, 6 (no markers). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

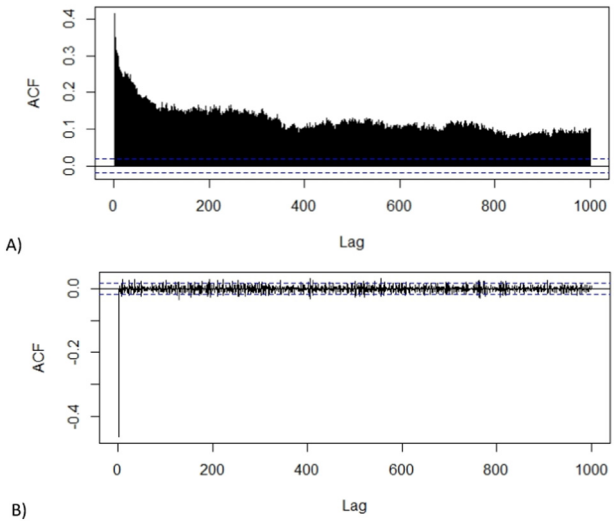

Fig. 9. Autocorrelation function (starting at lag 1) for the log-reposes (A) and their first difference (B).

equal numbers of observed reposes in each grouping interval, with no requirement for the intervals to have equal lengths. For example, for an eruption dataset containing 1000 explosion records, we may choose twenty cut points (τ_h in Eq. 5) by assigning the repose durations at each 5th percentile as the cut points. The more intervals that can be used in the model the better, with a limit determined largely by computational time. Here, we used 100 cut points.

The threshold parameters α_h correspond to the cut points of the grouped time bins, and we require that they satisfy the order constraint $\alpha_1 < \alpha_2 . . . < \alpha_m$. For identifiability, we set $\alpha_1 = 0$ if an intercept is incorporated in $X\beta$. The matrix $X = [1\ Y]$ is a vector of 1's for a model application that only considers the repose intervals, and augments this by a matrix Y of explosion variables when they are included, as described below. We see from Eq. (5) that if α_h is held constant, the probability that a repose is shorter than τ_h decreases as $X\beta$ increases, i.e., reposes become longer. The discrete hazard function (probability that the repose length is between τ_{h-1} and τ_h given it is $\geq \tau_{h-1}$) corresponding to Eq. (5) is $h(\tau) = \pi_h/(1 - \gamma_{h-1})$ for $\tau_{h-1} < \tau \leq \tau_h$.

This use of a non-parametric survival model is novel in volcanology, as the typical size of eruption records is insufficient. However, Smethurst et al. (2009) considered a piecewise linear non-parametric model for Mt Etna flank eruptions, with the eruptive probability remaining constant over partitions of the time history.

Fitting the model (5) to data is done via Markov Chain Monte Carlo (MCMC) estimation, as detailed in Goldstein et al. (2007) and Goldstein (2011): At any given iteration of the MCMC estimation algorithm, assume that we have current estimates of the $\alpha = \{\alpha_h\}$ parameters. We sample a value of the latent normal scale such that, if we observe category h $(1 < h < m)$ then we sample from the normal interval $[(\alpha_{h-1} - X\beta), (\alpha_h - X\beta)]$ with associated probability π_h. The updated β parameters β^* can then be sampled using a Gibbs step, assuming a diffuse prior with a resulting multivariate normal posterior. The updated α parameters (α^*) are derived from the α and these are sampled using a Metropolis step, accepting a new sample α^* with probability $\min(1, P_{\alpha^*}/P_\alpha)$, noting that the component of the likelihood associated with a particular ordered category is $P_\alpha = \prod_h \gamma_h$. Passarelli et al. (2010) have also used MCMC methods to fit hierarchical renewal models to repose data.

The advantages of this formulation over the parametric proportional hazard model are that 1) it makes no particular distributional assumption, which aligns with the different distribution fits to these data identified by Watt et al. (2007), 2) the coefficients are treated as random variables, in a fully Bayesian approach and hence epistemic variability is accounted for, and 3) the MCMC procedure automatically accounts for missing data through multiple imputation, which would have to be manually performed for parametric models using, for example, imputation or expectation-maximisation methods. For future applications, a multi-level formulation is potentially available to deal with different regimes.

For many of the variables, up to 60% (Fig. 3) of the data values are missing in the Sakura-jima dataset at various times. We assume that missing data are randomly distributed, or at least random conditional on the model parameters (Missing at Random, MAR). However, we acknowledge that missing information could be related to the subsequent repose duration (Fig. 4). If we were to remove records with missing variables this would substantially reduce the size of the dataset and the efficiency of our parameter estimates. Instead we can retain all the data and use an extension of the multiple imputation technique for missing values (Rubin, 1987). Goldstein et al. (2014) describes how the model

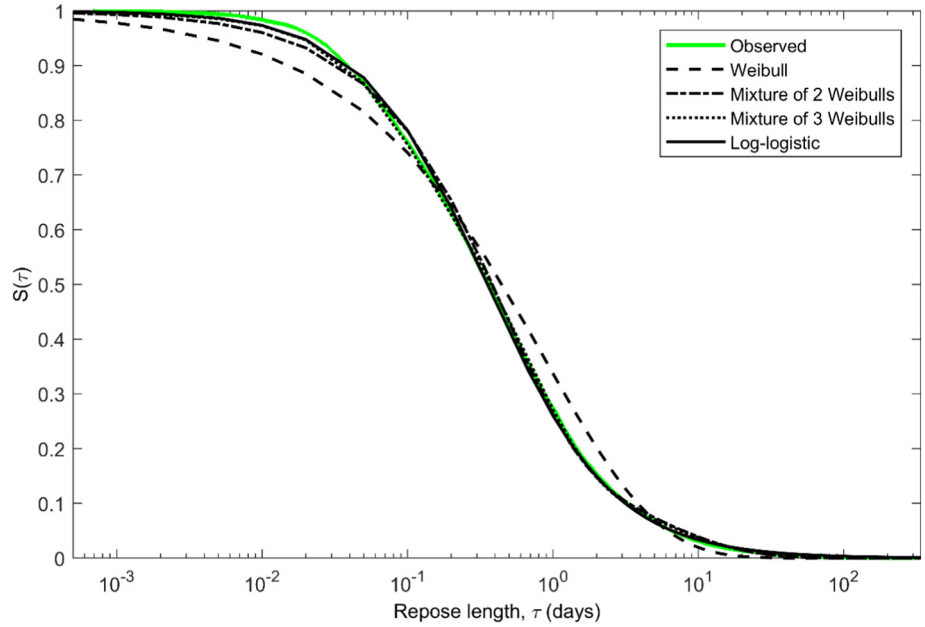

Fig. 10. Survival probabilities, $S(\tau) = P(\text{Repose} > \tau)$, for various renewal models, plotted against the observed 1956 to 2013 record.

can handle missing data in a fully efficient and fully Bayesian procedure that allows for missing data, where the predictor variables are either normal, e.g. log(plume height), or binary, e.g., plume colour, together with any interactions.

5. Model fitting and validation

Applying the model described in Section 4 to the 1956–2013 Sakurajima eruption record allows us to investigate repose intervals, and their relationship with explosion variables, without assuming a particular distribution. We can incorporate as much available data as possible, even with missing observations. Model outputs β and α_h characterise probability distributions, or survivor functions, for likely repose interval given a certain value of X, where X includes explosion variables, such as plume height or infrasonic amplitude.

The model was fitted to the entire data set, using stepwise removal of factors to obtain the best BIC, where BIC was estimated as

$$\mathrm{BIC} = \frac{N}{\sigma^2}\left(\mathrm{MSE} + \frac{p\ \log N}{N}\right)$$

(Hastie et al., 2001), where MSE is the mean of the square of the residuals, σ is the residual standard deviation of a complex model (all covariates included), N = number of data and p = number of parameters. The resulting model and parameter estimates are given in Table 2, with the posterior distributions for the β coefficients shown in Fig. 11, and their correlations in Table 3.

The plume size variables Q were excluded due to massive correlation with the plume height. The final model parameter estimates show that a higher plume indicates a longer subsequent repose, i.e. mean β is positive, effectively a time-predictable model (Bebbington, 2013b).

A black plume indicates a shorter subsequent repose, and a white plume a longer subsequent repose. A black plume contains the greatest concentration of ash, and so is possibly correlated with a larger mass of eruptible material in the conduit. Note that both categories are sparsely represented in the data set (Table 1), but it is nevertheless interesting that the model selected only the two 'extreme' plume colours (Fig. 6).

The Showa crater was inactive from the beginning of the data set, resuming activity in 2006 and being the sole crater producing explosions from March 2011. Regular explosions from Showa are recorded in our data set from 2008 and after this date we see a marked increase in the rate of eruptions, and hence a decrease in the repose lengths.

Although the estimated standard error for the coefficient of the infrasonic observations at station E exceeds its estimated mean, it was nevertheless significant by BIC measure. The infrasonic measurements (station E) were by far the shortest set of covariate data, available only between 1997 and 2000, and was marginally significant in Fig. 5. Its inclusion illustrates the power of the fully Bayesian imputation method to extract value from such scant data.

The serial autocorrelation incorporated through the previous reposes indicates that the activity level persists to some degree: short reposes tend to be followed by short reposes. The overall rate term represents regression towards a 'mean activity level', i.e. if the rate decreases, reposes become shorter, thus increasing the rate.

We see from Fig. 11 that most of the β distributions are reasonably symmetric and bell-shaped. The exceptions are the Showa crater and infrasonic coefficients. Occurrence of these factors is restricted to small parts of the dataset. Approximately 95% of the infrasonic values are imputed, which may explain the broader distribution of its coefficients. As seen in Table 3, the Showa coefficients are strongly correlated with all the other variables except plume colour. This relationship reflects increasing activity once the Showa crater reactivated.

The model was validated using the method in Bebbington (2013a). Each repose is predicted as a distribution, and we record the quantile of the predicted distribution in which the actual repose falls. In order to keep the computation within bounds, we adopted the compromise of selecting the model based on the full data set, and then using parameter estimates calculated from only the first 10,823 reposes without updating to forecast the final 1000 repose durations. We can compare the performance with the parametric model described by (2) and (4). Fig. 12 shows that the parametric model does not perform as well as the non-parametric Bayesian model (using the median estimates for the parameters) when using the derived test of uniformity at the 5% significance level.

Table 2
Non-parametric Bayesian survival model parameter estimates associated with the best BIC.

Factor	Mean β	Median β	SD β
Intercept	1.547	1.501	0.166
log of plume height	0.158	0.164	0.022
White plume	0.499	0.511	0.175
Black plume	−1.004	−1.061	0.500
Showa crater	−0.549	−0.551	0.028
log of infrasonic (E)	−0.024	−0.056	0.032
log of repose (lag 1)	0.151	0.152	0.006
log of repose (lag 2)	0.106	0.105	0.006
Rate of eruptions to date	0.674	0.677	0.156

6. Model forecasts

The purpose of the described modelling is to be able to forecast future repose intervals. In the model here, we use information from the most recent explosion, along with the previous two repose lengths as covariates to adjust the overall survival probability (i.e. the probability that the repose exceeds a given duration).

Fig. 13 shows how the forecast repose distribution varies as a function of selected preceding explosion variables. In order to isolate each factor, the other variables were held at their median values, and median β (Table 1) estimates were used. The observed distribution is also shown, as are the equivalent values from the parametric model.

Firstly, the median effects are almost identical under the two models (parametric versus non-parametric). These differ from the observed distribution due to the incorporation of non-linear effects and interactions of the explosion variables in the models. Secondly, we see that the Showa and log Repose-2 variables have a very similar range of effects under the non-parametric and parametric models. However, plume height has greater leverage in the parametric model, while the most recent repose has greater leverage in the non-parametric model. The repose is of course more accurately measured than the plume height, another point in favour of the non-parametric model.

7. Discussion and conclusions

In applying the non-parametric Bayesian survival model proposed here as a forecasting tool, no parametric distributional assumptions are made about repose intervals. The model can readily take advantage of whatever data are available, in whatever degree of completeness. Covariates such as explosion plume height or colour can be incorporated into the model. For Sakura-jima a number of explosion variables were found to provide statistically significant improvements in forecasts of likely future repose. Correlations between plume height and repose interval during similar explosions at Anak Krakatau in Indonesia have been found previously (Watt et al., 2007), but this is the first time that plume heights have been applied to quantitatively forecast explosion frequency.

The non-parametric model has the same limits on incorporation covariates (i.e. explosion variables) as in the proportional hazard model (Eq. 2) – only the last observation can be included. However, by modelling the threshold parameters as functions of explanatory variables, such as their moving average, the non-parametric model can be reformulated in a non-linear manner (Goldstein et al., 2014) to account

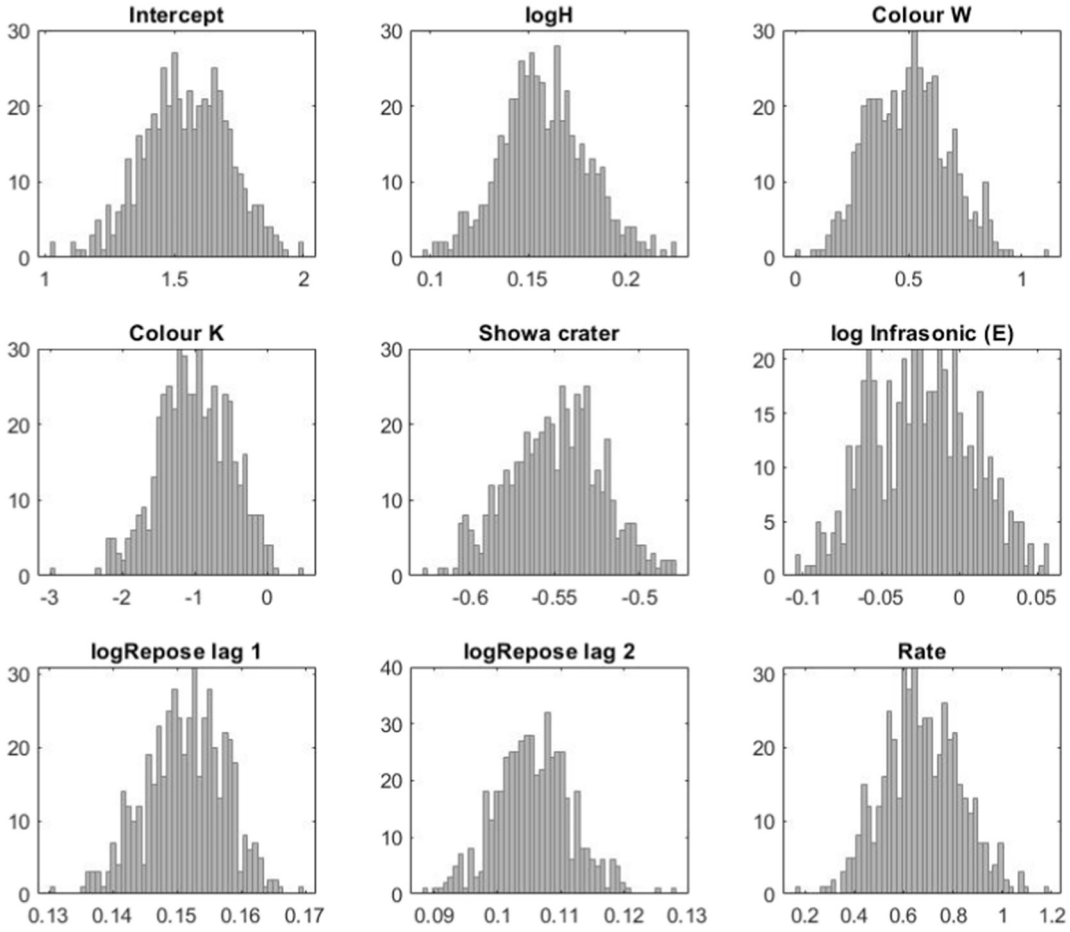

Fig. 11. Posterior distributions for β, using the non-parametric model. Colour K is black, W white.

for time-varying covariates (i.e. the entire history). This has been examined parametrically for eruptive volume (Bebbington, 2008), but not for other variables.

Earthquake types (e.g. b-type or hybrid) were not available to us. Incorporating them will require an elaboration of the model to account for time-varying (on an inter-explosion time scale) covariates. A simpler course would be to search for any seismic patterns that could be used over longer timescales (days to weeks) to provide an indication of future explosion frequency.

A further potential advantage of the model is the ability to account for missing data and multiple hierarchies in data (e.g. explosion sequences within eruption sequences within periods of activity), in addition to covariates that are predictive of repose intervals. The interactions of the 'Showa' term in the model is indicative of a system that may experience multiple regimes (Mulargia et al., 1987; Bebbington, 2007), where the statistical model shifts from one set of parameter values to another, due to a change in the underlying physical processes controlling the eruptions.

For illustrative purposes, we have used a least squares spline tool (freely available in MATLAB: SLM - Shape Language Modelling) to identify breakpoints within a linear regression, and thus divide the 1956 to 2013 eruption record into five broad periods (Fig. 14), with each period exhibiting a differing rate of activity:

- Period 1: January 1956 to July 1973 (n = 1699), in which Vulcanian explosions occurred on average once every four days (mean of 0.27 explosions/day); This could possibly have three sub periods;
- Period 2: August 1973 to November 1986 (n = 3648) was very active with nearly one explosion per day (mean of 0.75 explosions/day);
- Period 3: December 1986 to August 2001 (n = 2151) returned to a slower rate of activity with on average one explosion every two to three days (mean of 0.40 explosions/day).

Table 3
Correlations between multilevel model parameters. The lower left triangular portion of the table is the P values, upper the correlations (italicised if significant at 5%), W is a white plume, K a black.

	Int	logH	W	K	Showa	logInfE	logRep-1	logRep-2	Rate
Int		*−0.869*	0.004	−0.015	*−0.111*	*0.176*	−0.009	0.009	*−0.276*
logH	0.000		−0.001	−0.033	*0.353*	*0.114*	0.070	0.076	*−0.127*
W	0.931	0.976		0.007	0.039	0.036	−0.028	0.050	0.017
K	0.742	0.465	0.867		−0.070	−0.031	−0.017	−0.061	0.064
Showa	0.013	0.000	0.390	0.119		*0.184*	*0.246*	*0.230*	*−0.542*
logInfE	0.000	0.011	0.420	0.486	0.000		*0.090*	0.064	*−0.116*
logRep-1	0.837	0.117	0.526	0.698	0.000	0.043		*−0.270*	−0.015
logRep-2	0.844	0.089	0.264	0.176	0.000	0.156	0.000		*−0.096*
Rate	0.000	0.005	0.704	0.155	0.000	0.009	0.732	0.032	

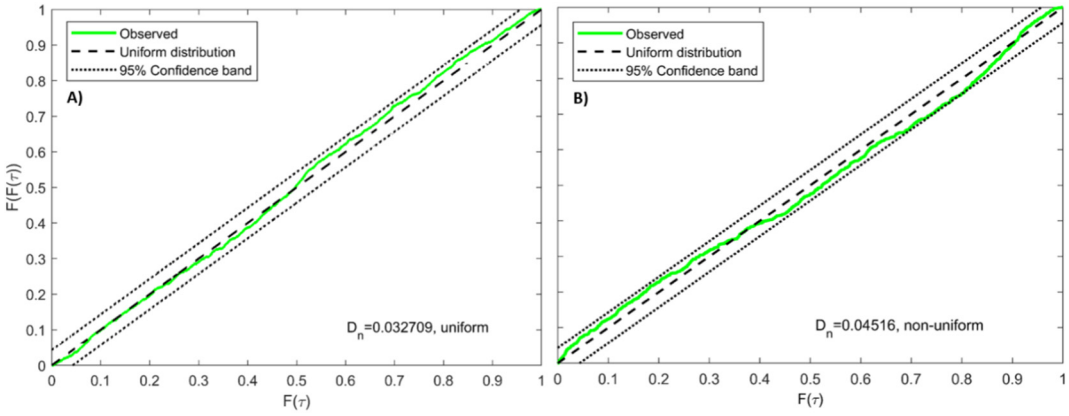

Fig. 12. Validation of (A) the non-parametric Bayesian model, and (B) the loglogistic proportional hazards model.

- Period 4: September 2001 to January 2009 (n = 174) represents a very quiet period of activity, with an explosion on average every two weeks (mean of 0.07 explosion/day).
- Period 5: February 2009 to December 2013 (n = 4156) was by far the most active period to date (mean of 2.35 explosions/day), and coincides with explosions mostly being generated from the Showa crater.

A multilevel model could be implemented by treating the 'period' as the second level of the hierarchy (Goldstein, 2011).

We can also consider multivariate models (Goldstein et al., 2009) where we may have several 'responses' in addition to the repose time. For example, the size of the explosion (e.g., plume height or seismic amplitude) could be modelled alongside the repose time, allowing an estimate of the correlation between them. This would augment the time-predictable model into a size and time-predictable model, previously only examined on an eruption level (Marzocchi and Zaccarelli, 2006; Bebbington, 2014).

The non-parametric Bayesian survival model may also prove useful for inferring the physical processes underlying explosions. The model selection process, suitably extended, can be used to test hypotheses about the significance of individual factors, and even factors acting in concert, uncomplicated by the assumption of a specific baseline distribution. There would thus be value in establishing if the non-parametric model developed here for Sakura-jima can be successfully applied to analogous open-system Vulcanian explosions for which we have incomplete or complete time-series of observations, e.g. Santa-

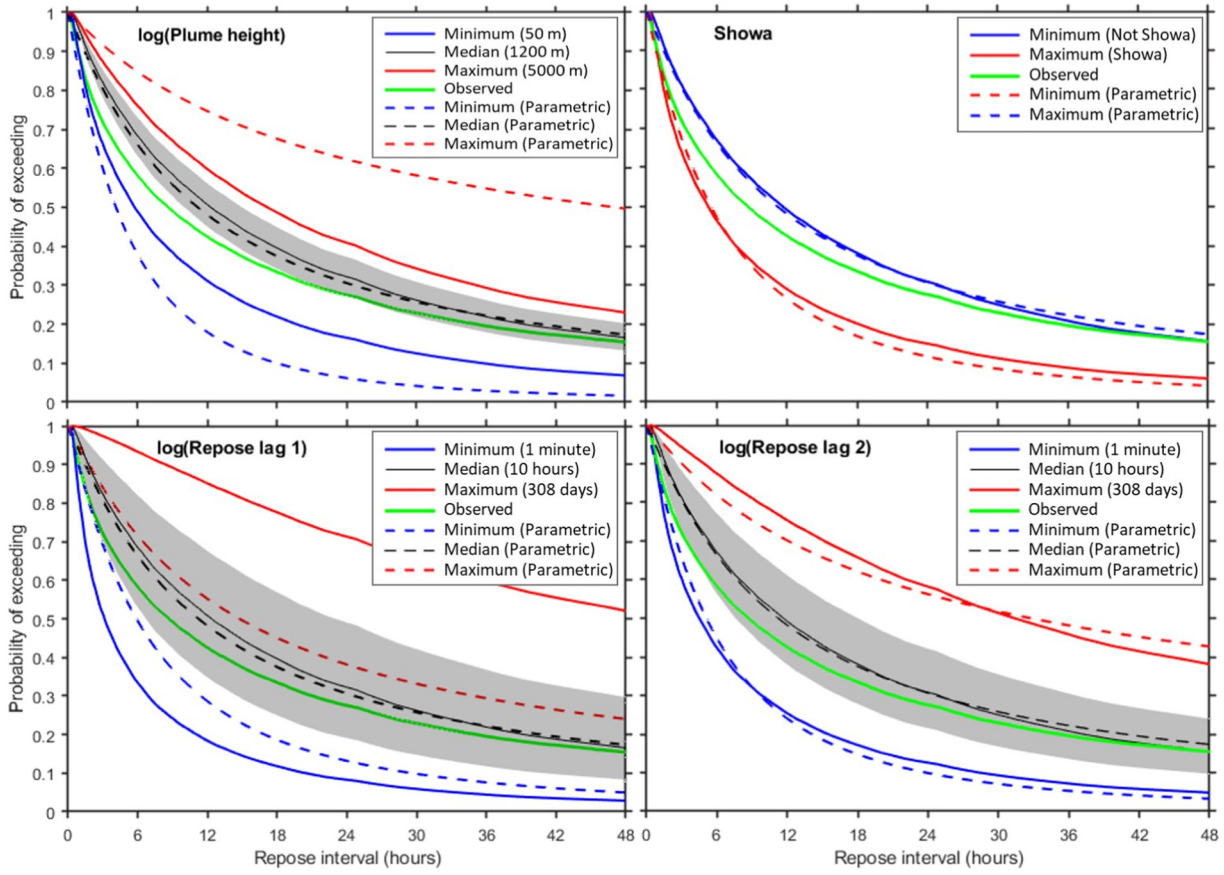

Fig. 13. Repose interval duration forecasts conditional on the variables from the last explosion being of a certain value. The grey area is the central 90% of the values; as a categorical variable, the central 90% of values for Showa cannot be represented.

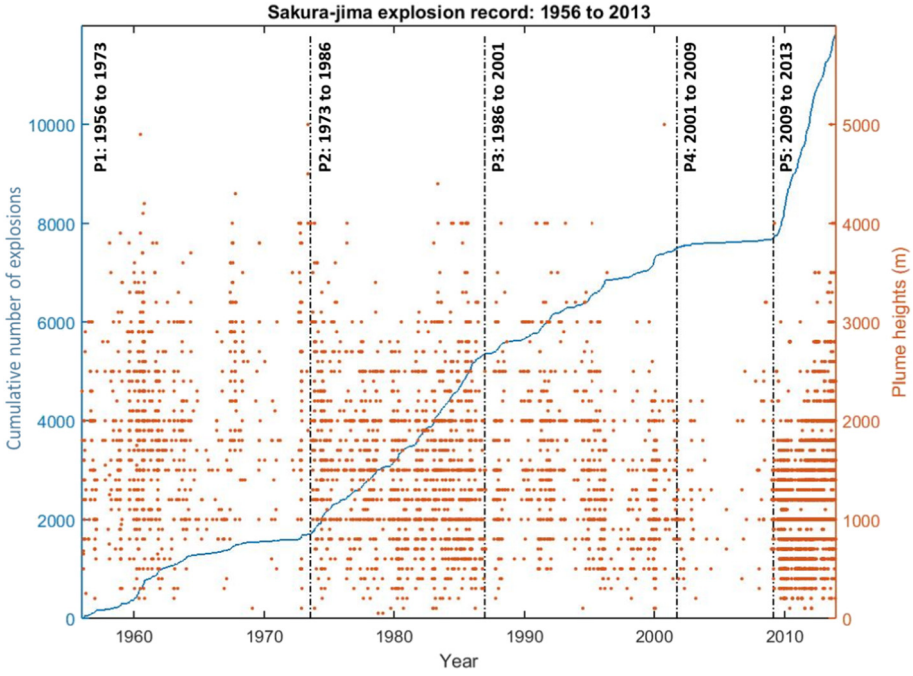

Fig. 14. Cumulative explosions with time and the recorded plume heights, where available, for those explosions. The 1956 to 2013 record is divided into five broad periods of similar activity.

Maria (1922), Popocatépetl (2005), Soufrière Hills (1997). If successfully applied, the model can be used to improve time-dependent hazard assessments for Vulcanian eruptions (e.g. Bonadonna et al., 2005; Jenkins et al., 2008; Jenkins et al., 2015).

However, the model is unable to provide viable results to forecast repose intervals between eruptions at closed-system volcanoes at longer timescales, due to insufficient data for a non-parametric approach. Here a parametric approach is required, unless the multi-level capability of the model (Goldstein et al., 2014) can allow for analogous volcanoes to be considered jointly to expand the available eruption dataset.

Acknowledgements

MB was supported by the New Zealand Natural Hazards Research Platform (Contract No. 2015-MAU-PC-01). RSJS acknowledges support from a Leverhulme Trust Emeritus Fellowship. This work comprises Earth Observatory of Singapore contribution no. 223. This research is supported by the National Research Foundation Singapore and the Singapore Ministry of Education under the Research Centres of Excellence initiative.

Appendix 1. Non-parametric Bayesian survival model glossary

The symbols used by the non-parametric Bayesian survival model, their description and use in the model.

Symbol	Description	Use in the model
τ_g	Time cut points. Often chosen as a percentile, so that data are equally divided between time bins	To provide time bins for the survival model
$\varphi(z)$	Standard normal distribution, with zero mean and standard deviation	The probit link function in the Bayesian formulation
α_g	Threshold parameter for a model with no covariates, i.e. explosion variables.	Model output for each time bin g
m	Number of time bins	
X	Explosion variable values	The survivor function is calculated for a given variable value, X. For example, the 5th, 50th or 95th percentile prior plume height.
β	Explosion variable coefficients	Model output for each explosion variable
α_k^*	Threshold parameter for a model with multiple covariates, for each time interval k	
q	Number of time-varying explosion variables	
π_g	The probability that an event occurs in time bin g	Calculated value from model outputs (Eq. 5)

Appendix 2. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jvolgeores.2019.04.008.

References

Aramaki, S., 1984. Formation of the Aira Caldera, Southern Kyushu, 22,000 years ago. J. Geophys. Res. 89 (B10), 8485–8501.
Bebbington, M., 2007. Identifying volcanic regimes using hidden Markov models. Geophys. J. Int. 171, 921–942.
Bebbington, M., 2008. Incorporating the eruptive history in a stochastic model for volcanic eruptions. J. Volcanol. Geotherm. Res. 175, 325–333.
Bebbington, M.S., 2013a. Assessing probabilistic forecasts of volcanic eruption onsets. Bull. Volcanol. 75, 783.
Bebbington, M.S., 2013b. Models for temporal volcanic hazard. Statistics in Volcanology 1, 1–24.
Bebbington, M.S., 2014. Long-term forecasting of volcanic explosivity. Geophys. J. Int. 197, 1500–1515.
Bebbington, M.S., Lai, C.D., 1996. On non-homogeneous models for volcanic eruptions. Math. Geol. 28, 585–600.
Bonadonna, C., Connor, C.B., Houghton, B.F., Connor, L., Byrne, M., Laing, A., Hincks, T.K., 2005. Probabilistic modeling of tephra dispersal: Hazard assessment of a multiphase rhyolitic eruption at Tarawera, New Zealand. J. Geophys. Res. 110 (B03203), 1–21.
Connor, C.B., Sparks, R.S.J., Mason, R.M., Bonadonna, C., Young, S.R., 2003. Exploring links between physical and probabilistic models of volcanic eruptions: the Soufrière Hills volcano, Montserrat. Geophys. Res. Lett. 30 (13), 34.
Goldstein, H., 2011. Multilevel Statistical Models. fourth edition. Wiley, Chichester.
Goldstein, H., Bonnet, G., Rocher, T., 2007. Multilevel structural equation models for the analysis of comparative data on educational performance. J. Educ. Behav. Stat. 32, 252–286.
Goldstein, H., Carpenter, J., Kenward, M.G., Levin, K.A., 2009. Multilevel models with multivariate mixed response types. Stat. Model. 9 (3), 173–197.

Goldstein, H., Carpenter, J.R., Browne, W.J., 2014. Fitting multilevel multivariate models with missing data in responses and covariates that may include interactions and non-linear terms. J. R. Stat. Soc. Ser. A (Stat. Soc.) 177 (2), 553–564.
Green, R., Bebbington, M.S., Cronin, S.J., Jones, G., 2013. Geochemical precursors for eruption repose length. Geophys. J. Int. 193, 855–873.
Hastie, T., Tibshirani, R., Friedman, J., 2001. Elements of Statistical Learning. Springer, New York (533 pp).
Jaquet, O., Sparks, R.S.J., Carniel, R., Mader, H.M., Coles, S.G., Connor, C.B., Connor, L.J., 2006. Magma memory recorded by statistics of volcanic explosions at the Soufrière Hills volcano, Montserrat. Statistics in Volcanology. Special Publications of IAVCEI. Geological Society, London, pp. 175–184.
Jenkins, S., Magill, C., McAneney, J., Hurst, A.W., 2008. Multi-stage volcanic events: tephra hazard simulations for the Okataina Volcanic Center, New Zealand. J. Geophys. Res. Earth Surf. 113 (F04012).
Jenkins, S.F., Barsotti, S., Hincks, T.K., Neri, A., Phillips, J.C., Sparks, R.S.J., Sheldrake, T., Vougioukalakis, G., 2015. Rapid emergency assessment of ash and gas hazard for future eruptions at Santorini Volcano, Greece. J. Appl. Volcanol. 4 (16), 1–22.
JMA, 1956a. Volcano Monthly Report: Distant View Observation Data. vol. 1 to 37. Japan Meteorological Agency Volcanological Bulletin (available on CD/DVD thereafter).
JMA, 1956b. Volcano monthly report: Volcanic earthquake data. Japan Meteorological Agency Seismological Bulletin. vol. 61 to 120 (available on CD/DVD thereafter).
JMA, 2015. http://www.data.jma.go.jp/svd/vois/data/tokyo/STOCK/bulletin/enbo.html, Accessed date: April 2016.
Kobayashi, T., Ishihara, K., Hirabayashi, J., 1988. A guide book for Sakurajima Volcano, Kagoshima, Japan.
Marzocchi, W., Bebbington, M., 2012. Probabilistic eruption forecasting at short and long time scales. Bull. Volcanol. 74 (8), 1777–1805.
Marzocchi, W., Zaccarelli, L., 2006. A quantitative model for the time-size distribution of eruptions. J. Geophys. Res. 111 (B4), B04204.
Mendoza-Rosas, A.T., De La Cruz-Reyna, S., 2010. Hazard estimates for El Chichon volcano, Chiapas, Mexico: a statistical approach for complex eruptive histories. Nat. Hazards Earth Syst. Sci. 10, 1159–1170.
Mulargia, F., Gasperini, P., Tinti, S., 1987. Identifying different regimes in eruptive activity: an application to Etna volcano. J. Volcanol. Geotherm. Res. 34, 89–106.
Ogata, Y., 1988. Statistical models for earthquake occurrences and residual analysis for point processes. J. Am. Stat. Assoc. 83, 9–27.
Passarelli, L., Sandri, L., Bonazzi, A., Marzoochi, W., 2010. Bayesian hierarchical time predictable model for earthquake occurrence: an application to Kilauea volcano. Geophys. J. Int. 181, 1525–1538.
Pyle, D.M., 1998. Forecasting sizes and repose times of future extreme volcanic events. Geology 26 (4), 367–370.
Rubin, D.B., 1987. Multiple Imputation for Non-response in Surveys. Wiley, Chichester.
Smethurst, L., James, M.R., Pinkerton, H., Tawn, J.A., 2009. A statisical analysis of eruptive activity on Mount Etna, Sicily. Geophys. J. Int. 179, 655–666.
Turner, M., Cronin, S.J., Bebbington, M., Platz, T., 2008. Developing a probabilistic eruption forecast for dormant volcanos; a case study from Mt Taranaki, New Zealand. Bull. Volcanol. 70, 507–515.
Turner, M.B., Cronin, S.J., Bebbington, M., Smith, I.E.M., Stewart, R.B., 2011. Relating magma composition to eruption variability at andesitic volcanoes: a case study from Mount Taranaki, New Zealand. GSA Bull. 123, 2005–2015.
Udagawa, S., Imanaka, H., Koyaguchi, T., Takayasu, H., 1999. Statistical analysis of occurrence of eruptions at Sakurajima Volcano. Proceedings of the Volcanological Society of Japan. vol. 32 (In Japanese).
Watt, S.F.L., Mather, T.A., Pyle, D.M., 2007. Vulcanian explosion cycles: patterns and predictability. Geology 35 (9), 839–842.
Yokoo, A., Iguchi, M., Tameguri, T., Yamamoto, T., 2013. Processes prior to Outbursts of Vulcanian Eruption at Showa Crater of Sakurajima Volcano. Bull. Volc. Soc. Japan 58 (1), 163–181.